2005 IEEE International Conference on Systems, Man and Cybernetics
Waikoloa, Hawaii October 10-12, 2005

Fault Management during Dynamic Spacecraft Flight: Effects of Cockpit Display Format and Workload

Valerie A. Huemer
San Jose State University Foundation
NASA Ames Research Center
Moffett Field, CA, U.S.A.
Valerie.A.Huemer@nasa.gov

Michael P. Matessa, Robert S. McCann
Human Factors Research and Technology Division
NASA Ames Research Center
Moffett Field, CA, U.S.A.
{Michael.P.Matessa, Robert.S.McCann}@nasa.gov

Abstract - *A proposed redesign of the shuttle cockpit display formats improves the correspondence between system summary displays and crewmembers' mental models of systems architecture and functional mode. We report the results of a part-task simulation that assessed the impact of the redesigned displays on participants' ability to perform various steps in the process of diagnosing and recovering from systems malfunctions. Participants were airline pilots who received a modest amount of training on the tasks required of shuttle crews during nominal and off-nominal ascents. With respect to fault management performance, both errors of omission and commission were reduced with the redesigned displays. Fault management errors were further categorized within a cognitive-stage information processing framework. Error rates increased steadily from early to late stages of processing, but more so for the current displays than for the redesigned displays. We conclude that classifying and analyzing errors made by participants with relatively low levels of training provides a useful methodology for assessing and evaluating human-centered design modifications to spacecraft displays.*

Keywords: Errors, omissions, commissions, information processing model, spacecraft, space shuttle, MEDS, CAU.

1 Introduction

The space shuttles flying today were designed and built over 25 years ago. Despite several hardware improvements since that time, including conversion to a "glass cockpit" in 2000, crew workload remains high, particularly in off-nominal situations. Cockpit automation to support the crew in off-nominal conditions is minimal, leaving most emergency operations to be performed by the crew and mission control personnel. For example, the cockpit caution and warning system automatically annunciates out-of-limits sensor readings that accompany systems malfunctions. However, it is up to the crew and mission control to jointly diagnose the root cause of the anomalous reading(s), access the appropriate remedial actions in paper flight data files (FDFs), perform the specified actions, and verify their effectiveness. For many malfunctions, these activities increase crew workload to the point where crewmembers have very little margin to deal with any additional problems.

Another factor contributing to crew workload is that the cockpit displays were designed to accommodate 1970's era limitations in electronic display technology and onboard computing capability [1]. Consequently, displays are data-source-oriented (rather than task-oriented), so the crew is often required to navigate through several displays to gather the information needed to complete a task. The displays are often poorly organized and highly cluttered, taking the form of closely-spaced tables of alphanumeric data that require considerable mental translation in order to infer the current operational status or functional mode of the onboard systems.

To address the shortcomings of the current shuttle cockpit display formats (collectively known as the "Multifunction Electronic Display System" [MEDS]), the current display formats have been completely redesigned as part of a recently completed Cockpit Avionics Upgrade (CAU) project (unfortunately, due to budget considerations, the display redesigns are not going to be actually implemented on the shuttles before they are retired in 2010). One of the fundamental human factors principles guiding the redesign of the systems summary displays (those displays that provide information about systems status and system functioning) was that these displays should be "transparent" to the actual working system, so that the operator can "see through" the displays to "what is going on" with the underlying system [2]. The resulting display formats, described by McCandless and McCann [3], consolidate systems information onto single, task-oriented systems summary displays to reduce the need for display navigation. Many of these display formats incorporate dynamic graphical depictions of systems components, providing "at a glance" indications of the operational status and functional mode of the subsystem. The following illustration shows how these human factors principles were applied in the upgraded cockpit.

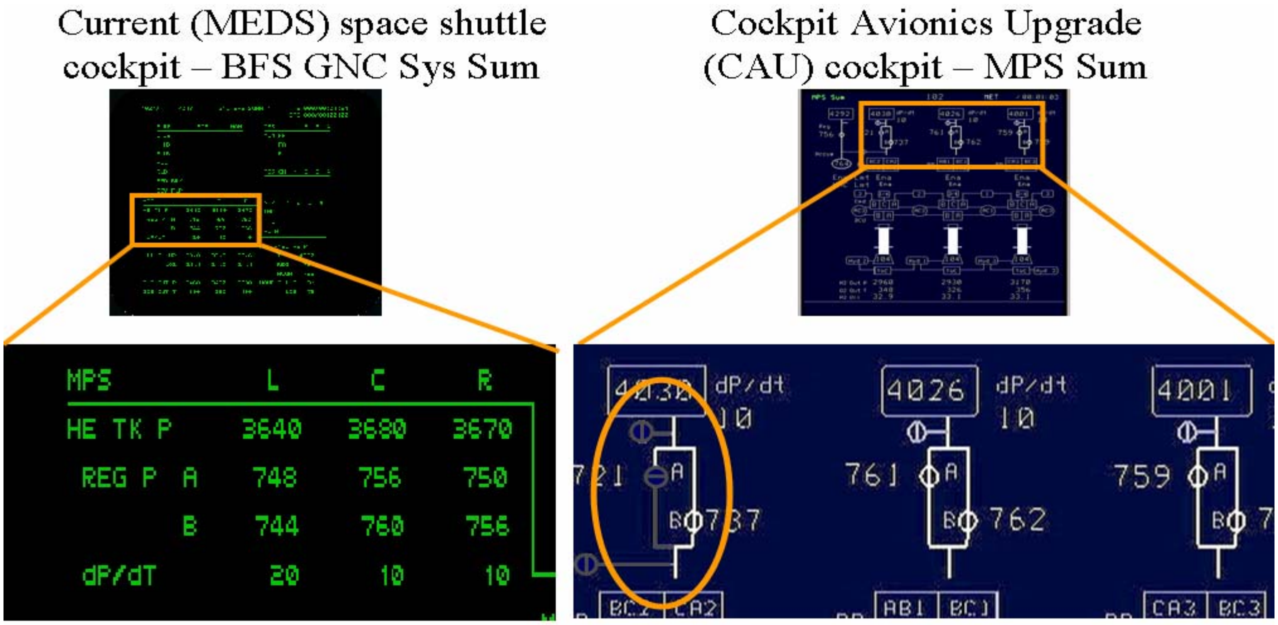

Figure 1. MEDS vs. CAU Cockpit Displays

1.1 An Illustration of Human Factors Concepts in the CAU Redesign

Consider how information is displayed to the crew about the helium supply subsystems that play a key role in the safe operation of each of the shuttle's three main engines. Each engine requires a constant supply of helium to pressurize an intermediate seal in the high pressure oxidizer turbopump. Helium is supplied from three sets of tanks, one set for each engine. Helium from each tank set flows to the engine through two redundant feedlines (legs), designated "A" and "B". Each leg has its own regulator that reduces the pressure coming from the tanks, and its own isolation valve that can be closed to stop the helium flow through that leg (these valves are normally open).

The left side of Figure 1 is the "BFS GNC Sys Sum 1" display from the current MEDS space shuttle cockpit (in reality, each cockpit display measures about 7.5 in x 7.5 in) with the enlarged area showing how information about the workings of each helium supply system is displayed to the crew. As shown in the figure, this information takes the form of a matrix of digital values indicating the tank pressures, the regulated pressures through legs A and B, and the change in tank pressures over time (dp/dT; a measure of the rate at which helium is being supplied to the engine).

The right side of Figure 1 shows the upgraded CAU main propulsion system (MPS) summary display (MPS Sum); the enlarged area shows the region devoted to the helium supply systems. On this display, the alphanumeric matrix from the current display (BFS GNC Sys Sum 1) is replaced with a graphical representation, showing flow (in the form of bright white lines) and circles representing the left(A) and right(B) leg isolation valves.

These graphical elements are dynamic, changing to depict the current state of the system. For example, as shown in the circled area of Figure 1, a crewmember has closed the helium isolation valve for leg A on the left engine. The section of the flow line inside the valve indicator (i.e., inside the circle) is rotated to the point where it is perpendicular with respect to the line, indicating that the valve is now in a closed position. In addition, the loss of flow through the affected leg (A) is signaled by the corresponding flow line turning from bright white to dark grey. In this way, the upgraded display provides a more direct and presumably more intuitive representation of the underlying system (its current state and operational mode) than the BFS GNC Sys Sum 1 (current) display.

Our interest is in how these display upgrades (and the corresponding changes in training techniques) affect fault management under various levels of workload. In the next section, we develop an empirical framework to investigate this issue.

1.2 Errors and Information Processing

The primary purpose of cockpit systems summary displays is to assist crewmembers in maintaining situational awareness of systems health and functioning, and to handle systems malfunctions (faults) in an accurate and timely manner. Fault management is a very complex process, requiring many constituent skills and activities, creating multiple opportunities for generating different types and forms of operator error. For their part, astronauts receive an extraordinary amount of simulator training to master these skills and be able to perform to a high level of accuracy in an emergency. These very highly-trained individuals should obviously play an important role in evaluating new spacecraft display design concepts. However, there is a drawback in using astronauts as the only source of evaluation: they are so highly trained that they make few errors.

An error analysis is useful in many ways. It provides clear guidance as to what task components are most difficult, and what features of a display redesign might provide the greatest performance benefits early in training. This information can be used to improve and shorten training programs. Second, by identifying which constituent activities and processes are most error-prone, we can identify which activities are the most attractive candidates for automation in next-generation spacecraft. Finally, by observing what kind of errors are committed and when they occur, we can infer what mental operation was being attempted at the time of the error. This information can assist with the development of human performance models of the fault management process. Such models can then be used to predict the effectiveness of candidate operational concepts for vehicle health management in next-generation vehicles.

Accordingly, this paper reports the results of a CAU redesign evaluation using a sample of commercial airline pilots who received only a modest amount of training on shuttle cockpit operations and procedures. These individuals were highly motivated and experienced, and performed accurately enough to give us confidence that our training procedures included the requisite skills and procedures. At the same time, enough errors were committed to support meaningful analyses.

Many distinctions and categorizations of human error have been proposed by various investigators [4]. One common distinction is an error of *omission* versus an error of *commission*. An error of omission occurs when an operator fails to perform a certain action. An error of commission occurs when an operator performs an incorrect action. Conveniently, each spacecraft malfunction is associated with a very specific set of procedures, typically involving switch throws that reconfigure the affected system in order to exploit built-in operational redundancies. These actions are specified on cue cards or in flight data files (FDFs) that crewmembers access once the malfunction has been identified. Thus, we can explicitly associate errors of omission with failures to perform the activities identified in the FDFs, and errors of commission with any actions (i.e., incorrect or additional switch throws) not specified in the files.

The omission/commission classification is broad enough to neatly feed into other established error models, such as Reason's "slips, lapses, and mistakes" model and Rasmussen and Jensen's "skill-, rule-, and knowledge-based performance" model [5]. However, classification of errors into those of omission and commission is only a preliminary step towards developing a comprehensive model of fault management performance in a spacecraft cockpit. Errors can be further distinguished by the context of the environment, what the intention of the operator was, and whether the error occurred during perception or execution of a task [6]. For example, Wiegmann and Shappell [7] tested the utility of a traditional information processing model, proposed by Wickens and Flach [8], to categorize factors involved in U.S. Navy and Marine aviation accidents between 1977 and 1992. The model assumes that when performing a cockpit task, the operator's information processing progresses through a series of stages or mental operations between stimulus input and response execution, namely: 1) "Sense," 2) "Attend," 3) "Recognize," 4) "Decide," and 5) "Execute."

In the first, "Sense" stage, features of stimuli are stored temporarily in perceptual buffers. In the second, "Attend," stage, attentional orienting operations influence what features and information are processed in all of the later stages. In the "Recognize" stage, these features are integrated into meaningful elements and identified. In the "Decide" stage, choices are made about how to react to the information. Finally, in the "Execute" stage, the information is decomposed into the cognitive and motor activities required to make the correct response.

If errors are to be used to inform future improvements in spacecraft design, an approach that associates errors with specific information processing stages may be more informative than a simple omission/commission classification. For example, errors of omission would have different design implications if they occurred as a result of failing to notice the existence of the malfunction entirely, as opposed to failing to follow the procedural steps defined in the FDFs. In the first case, more effective alarms would be an obvious design solution. In the second, an improved, possibly electronic, version of the FDF might be warranted.

Thus, as Wiegmann and Shappell did for aviation, we can use the information processing stages identified by Wickens and Flach to gain insight into which cognitive activities involved in spacecraft fault management produce the most errors. Within this framework, both errors of omission and errors of commission could arise in any processing stage. Examples of these errors would be recognition errors due to inaccurate decoding of the fault message, decision errors by deciding to use the wrong procedure, and execution errors by omitting a step in the procedure.

We can illustrate this classification system with respect to the helium system described earlier. Suppose a "Leg A" regulator failure in the left engine helium supply system occurs, which produces an off-nominal high rate of flow of helium from the Left Engine supply tanks. In this example, the "Sense" and "Attend" stages are associated with perceiving the cockpit alarm, perceiving the off-nominal indications on the relevant cockpit displays (BFS GNC Sys Sum 1 in the MEDS cockpit, MPS Sum in the CAU cockpit), and directing visual attention to the associated fault message (written as "MPS He P" in both cockpits). The "Recognize" stage corresponds to the crew member reading the fault message, associating it with the off-nominal indications on the appropriate display, understanding that the problem is a regulator failure in the left engine helium supply system, and how urgent the problem is. The "Decide" stage is associated with accessing and understanding the appropriate set of procedures in the FDF. Finally, in the "Execute" stage, the instructions in the FDF are decomposed into the cognitive and motor activities required to locate the appropriate switches and make the correct switch throws.

Wiegmann and Shappell found that the frequency of errors in aviation accidents increased monotonically with the stage of information processing. For example, they found that errors in the final (Execute) stage were most frequent (45.48%), with successively fewer errors in the earlier Decide stage (29.54%), in the Recognize stage (14.87%), in Attend operators (7.28%), and in the Sense stage (2.84%). If there are similar error patterns in spacecraft fault management, then error-reducing technologies designed for aviation may be useful for space, and vice-versa.

1.3 Current Study

In this article, we report a portion of the results of an empirical assessment of the impact of the upgraded display format redesigns (CAU), along with associated changes in training techniques, on the fault management performance of airline transport pilots. We also wanted to assess whether the impact of the redesigns varied with workload; thus, we included a low workload condition, in which only one malfunction had to be managed, and a higher workload condition in which multiple malfunctions had to be managed. Empirically, we looked at procedural accuracy and errors of omission and commission with each display format, as well as in what information processing stage these errors occurred. In this way, we explored how errors in each cognitive process are affected by display design and changes in workload.

As we noted, astronauts are extremely highly trained on fault management procedures, to the point where they make very few outright procedural errors. However, it is possible, under high stress or high workload, that errors can creep in, even for experts. Errors committed by individuals with less spacecraft-specific training can provide insight into the error patterns astronauts may have during these unusual conditions. Thus, in our study, we recruited highly-experienced airline transport pilots, trained them on the basic information processing requirements in the space shuttle cockpit, and observed the errors that these "non-astronauts" made in operating the spacecraft. Experienced airline pilots were chosen because they already had familiarity with flight dynamics and effective aviation scanning techniques.

2 Method

2.1 Participants

Twelve airline transport pilots, with an average of 16,800 flight-hours on various aircraft, participated in our experiment. Six pilots were assigned to the "current display" group (the MEDS cockpit) and the remainder to the "upgraded display" group (the CAU cockpit).

2.2 Apparatus

The study was conducted in a single-person part-task spacecraft cockpit simulator at the Intelligent Spacecraft Interface Systems (ISIS) laboratory at NASA Ames Research Center. The simulator currently emulates key cockpit features (displays and switch panels) accessible to the left-seat crewmember (i.e., the Commander) in the shuttle cockpit. The simulator is housed in a structure that supports twelve 20" touch-panel liquid crystal display (LCD) monitors. Four monitors in the direct forward field of view were used to represent the forward cockpit display formats. Seven monitors were used to represent the side and overhead panels, and a 12" monitor represented the keyboard. Touch-panel LCD monitors are used to allow the subject to manipulate switches, as required. The monitors and an audio system (to provide engine noise and alarm annunciation) are controlled by a multi-platform (SGI and PC) computer network.

2.3 Procedure

2.3.1 Single Malfunction (Low-workload) Testing

Prior to testing, each group of pilots ("Current Display Group" and "Upgraded Display" Group) participated in a week-long training course that covered basic shuttle systems, ascent-related displays, display navigation (i.e., keyboard) functions, nominal display monitoring requirements during ascent, and procedures for working several possible malfunctions. Each pilot was also given a two-hour familiarization session in the simulator prior to testing.

An important point to note is that, while each group received the same *amount* of training (from a temporal perspective), the content of the training differed, as it was customized to fit the display environment. That is, the Current Display Group was familiarized with the MEDS displays, whereas the Upgraded Display Group was familiarized with the CAU displays. However, since the predominantly-textual MEDS displays were poorly designed to support development of an understanding of the underlying systems, the current display group was trained on systems architectures using graphical depictions of the systems derived from materials used for astronaut training at Johnson Space Center. Since the CAU displays are graphical and dynamic, they provide much better support for developing systems understanding. Consequently, the actual CAU cockpit displays played a more prominent role in training the Upgraded Display Group to understand the functioning of the systems than the MEDS displays did for the Current Display Group. Thus, the Upgraded Display Group had the advantage of learning about the underlying systems with the actual displays used in testing.

Participants in each group then completed simulated shuttle missions from launch to main engine cut-off (MECO), lasting 8.5 minutes of simulator ("Mission-elapsed") time (MET). Eight runs were nominal (no malfunctions) and four were off-nominal. Nominal runs were included in all conditions of our experimental design to provide an environment in which malfunctions were not expected on every run.

During nominal trials, participants performed several mandatory navigational and systems parameter checks, such as solid-rocket booster separation at 2:00 MET, functional status of the Freon system at 3:00 MET, and the vehicle roll to heads-up attitude at 5:40 MET.

On each off-nominal run, one of four systems malfunctions was inserted: 1) a malfunction involving a regulator in the helium supply subsystem for one of the shuttle's three main engines (annunciated to the participants as a main engine helium pressure problem); 2) a leak in the external tank holding the fuel (liquid hydrogen) for the main engines (annunciated to the participants as a low ullage pressure problem); 3) a failure of one of the four onboard general purpose computers (GPC) to maintain synchronous operations with the remaining three machines (annunciated to the participants as a GPC fail-to-synch problem, or 4) a failure in the vehicle's thermal management system responsible for cooling the freon loops during ascent (annunciated to the participants as an "Evaporator Out Temperature High" problem). Nominal and off-nominal trials were randomly interspersed with the following restrictions: 1) the first and last trials were always nominal, and 2) no more than two malfunction trials in a row could occur.

2.3.2 Multiple-Malfunction (High-Workload) Testing

Approximately four months after completing the single-malfunction testing, both groups of participants returned for a one-day refresher course on basic shuttle systems, nominal monitoring tasks during ascent and resolution procedures for specific possible malfunctions. Each pilot was given one nominal trial immediately before testing so that they could familiarize themselves with the simulator cockpit layout and nominal monitoring tasks.

Participants in each group then completed four additional ascent runs: a nominal run, followed by two off-nominal runs with three malfunctions inserted during each, and a final nominal run. The specific malfunctions (described in the previous section) were combined as follows:

1) In the "HGE" scenario, the helium failure occurred at 1:50 MET, the GPC failure occurred at 2:00 MET, and the Evaporator failure occurred at 3:05 MET.
2) In the "GUE" scenario, the GPC failed at 1:50 MET, the Ullage pressure problem was inserted at 2:00 MET, and the Evaporator failure was inserted at 3:05 MET

Within each group, half (3) of the participants were randomly assigned to receive the "HGE" scenario during the first malfunction run (trial 2) and the "GUE" scenario on the next run; the other half of the group received the opposite order.

Participants were familiarized with each potential fault during training and simulator familiarization. They were instructed in the proper procedures for resolving each of the malfunctions, and where to access these procedures in an FDF. Simulator parameters and switch throws were recorded during each trial. After each trial, each participant answered questionnaires rating workload and situational awareness.

3 Results

3.1 Accuracy

In order to compare error rates across malfunctions, it is important to equalize the number of fault management steps required for resolution of each type of malfunction. Thus, for error analyses, we included only those systems malfunctions that required two discrete actions (switch throws) to complete the procedure. This led to the exclusion of the ullage pressure malfunction, which only required one switch throw.

A "correct" resolution of a malfunction was defined using the conservative criteria of both actions being executed, in the correct order, with no errors of commission. The proportion of malfunction procedures performed correctly (of the total number of procedures for each subject) was calculated and used for comparison of performance between the two groups. Performance measures were collapsed over the different types of malfunction scenarios.

A split-plot ANOVA including Cockpit Display Format (Current Display Group versus Upgraded Display Group) and Workload (single versus multiple malfunction trials) as factors revealed a significant main effect of Cockpit Display Format ($F(1,20)=5.93$, $p<0.05$), with the Upgraded Display Group resolving considerably more malfunctions correctly (68%) than the Current Display Group (36%). There was no significant main effect of Workload and no significant interaction. However, there was a trend toward higher accuracy in the low workload condition. The Current Display Group resolved 39% of the malfunctions correctly and the Upgraded Display Group resolved 72% of malfunctions correctly under low workload. Under high workload, the Current Display Group performed 33% of the procedures correctly and the Upgraded Display Group performed 63% correctly.

3.2 Errors of Omission versus Commission

Errors of omission and errors of commission were tabulated for each trial. The average number of errors (omission or commission) per off-nominal run (collapsed across type of malfunction) are shown in Figure 2. A split-plot ANOVA, including the factors Cockpit Display Format (Current Display Group versus Upgraded Display Group), Error Type (Omission versus Commission) and Workload (single malfunction runs versus multiple malfunction runs) revealed a significant effect of Cockpit Display Format ($F(1, 40) = 4.95$, $p < 0.05$), reflecting the fact that more omission and commission errors were committed by the Current Display Group than by the Upgraded Display Group. There was also a significant effect of Error Type, $F(1, 40) = 4.11$, $p < 0.05$, with more errors of commission than errors of omission. There were no significant interactions.

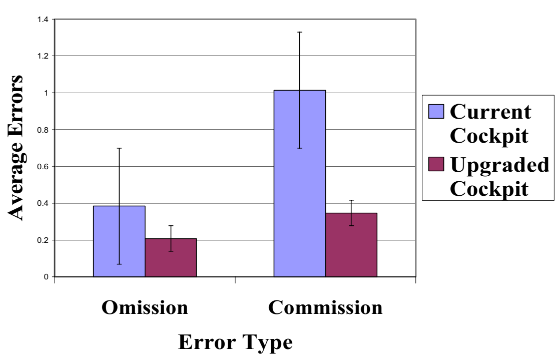

Figure 2. Average Errors per Off-Nominal Trial

3.3 Stage-Related Analyses

Wiegmann and Shappell [7] found that later stages in the information processing sequence were associated with more errors than earlier stages. We can see how this pattern relates to fault-management errors in a spacecraft by analyzing shuttle simulator errors with the information processing framework proposed by Wickens and Flach [8]. In the shuttle simulator experiments, some participants made multiple errors in a single run. On a real mission, the first error would cause a mishap, so we categorized the first error of an incorrect trial with the information processing framework and used those counts in the following analyses.

Every malfunction elicited some response from the pilots (either by manual acknowledgement or visual search of relevant regions of interest), so no errors were associated with the Sense or Attend stages. Errors in the Recognize stage include mistaking “Aff” (meaning affected) for “All” in the written procedures.[1] Errors in the Decide stage include not deciding on any procedure and performing a procedure associated with another malfunction. Errors in the Execute stage include omitting procedural actions and performing actions inappropriate for any currently active malfunction.

The proportion of errors in each stage was calculated by dividing the number of errors for each stage by the total number of errors across all stages. As in the Wiegmann and Shappell aviation study, errors monotonically increased with stage. Specifically, errors in the final (Execute) stage were most frequent (70.59%), followed by successively fewer errors in the earlier Decide stage (19.61%), the Recognize stage (9.8%), and finally 0% in both the Sense and Attend stages.

The stages of Wickens and Flach’s information processing model can also be used to examine changes in error patterns between current display and upgraded display participants. Figure 4 shows the average number of errors per person in each information processing stage for the two display groups. Error bars in the figure represent standard error.

[1] This particular error, which was common during the single malfunction trials, was not included in the previous analysis because training during the familiarization sessions for the multiple-malfunction trials explicitly addressed the error. It is included here because it is an easily-understood Recognize error and because it does not affect later current vs. upgraded display analyses.

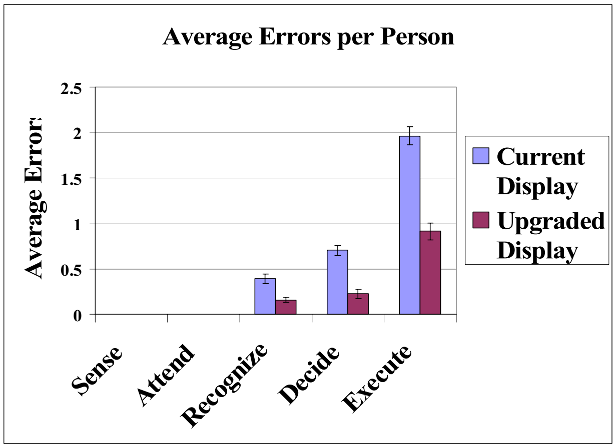

Figure 4. Errors per Person in Information Processing Stages and Cockpit Display Group

The error counts were analyzed using a 3-way ANOVA, including the factors Cockpit Display Formats (Current Display Group versus Upgraded Display Group),Workload (single malfunction runs versus multiple malfunction runs), and Information Processing Stage. There was no main effect of workload, and no significant interactions involving workload. There was a significant effect of Cockpit Display Format, F (1, 100) = 10.88, $p <$ 0.01, reflecting the fact that more errors were committed by the Current Display Group than the Upgraded Display Group. There was also a significant effect of Information Processing Stage, F (4, 100) = 23.17, $p < 0.001$, reflecting the monotonic relation between stage and proportion of total errors noted earlier. There was also a significant interaction of Cockpit Display Type and Information Processing Stage, F (4,100) = 3.48, $p < 0.05$, showing that as the Information Processing Stages progressed, errors increased at a greater rate in the current cockpit than in the upgraded cockpit. Although errors are reduced from the current displays (MEDS) to the upgraded displays (CAU) in the Recognize, Decide, and Execute stages, individual comparisons revealed that the reduction was significant only in the Decide stage ($t(10)$ = 2.19, $p<0.05$) and the Execute stage ($t(10)$ = 2.59, $p<0.05$).

4 Discussion

Analyses of overall malfunction-handling accuracy revealed significant improvements with the upgraded displays. An analysis of errors by information processing stage showed a significant decrease in errors from the current cockpit to the upgraded cockpit. This decrease was most pronounced for "Decide" and especially "Execute" errors.

Why did the upgraded CAU display provide more benefit to the final stages, particularly the execute stage? There are a number of possibilities. We believe the disproportionate benefit to the execute stage is due in part to better presentation of system state, and in particular better and more direct feedback concerning changes in state due to pilot action. For example, solving the regulator malfunction in the helium supply system requires two switch throws, one to stop the flow of helium through the leg containing the failed regulator and one to open an interconnect manifold to supplement the flow of helium into the affected engine from a backup helium supply system. On the BFS Sys Sum 1 display, (Current Display Group), the only feedback after taking the first action (closing Helium Isolation "A" valve) is a subtle change to the relevant digital pressure value, which may not have been sufficient to maintain awareness of the current configuration of the helium supply subsystem (i.e., that "Isol A" valve was closed, and there was no flow through leg "A"). This lack of awareness could have lead to the pilot performing an incorrect switch throw, and proceeding to the next step unaware of his error (viz., one pilot in the Current Display Group closed the isolation valve for the wrong leg [leg B], then performed the subsequent step, having no feedback that he had opened the wrong isolation valve).

The lack of situation awareness regarding system state associated with the current displays may also lead the pilot to lose his place in the procedure and skip a step. With the upgraded displays there is an update of a graphical representation showing helium flow, which reduces the chance of system state confusion and gives the pilot a better idea of where he is in the procedure and what steps still need to be done.

But the benefit of the upgraded displays may not be due solely to the addition of graphical feedback during the malfunction handling operation, since a novice looking at the graphics would probably not be able to understand them. As we noted, the display graphics are detailed enough that they were used to teach systems architecture and system functioning during training. This training presumably linked the pilots' mental model of system state and functioning to the same graphical representation that they used to work the malfunction procedures during the simulations. It is not clear whether the upgraded displays themselves, or the correspondence between the upgraded displays and the mental models developed during training, yielded the performance benefits observed in our study. If training turned out to be the critical factor, this would suggest that designers of systems summary displays for next-generation crewed spacecraft should adopt the graphical system summary design guidelines followed by the CAU project, and the resulting displays should play a central role in astronaut training programs.

Another interesting question is how the graphics and increased information in the upgraded displays (showing organization, system state, mode, etc.) affected mental workload. With more information to monitor, mental workload may actually increase. Future studies looking at eye movements could provide information on the effect of the upgraded graphics on workload and monitoring strategies.

The pattern of errors in the information processing model analysis may be used to inform future improvements to space cockpit design. For example, a new system to assist the pilot in the execution of procedures may be chosen over a new system to alert the pilot to a malfunction, since there were many more demonstrated errors in the Execute stage than in the Sense and Attend stages. Or perhaps execution stage activities should be automated, thus reducing the chance for errors in that stage. We are particularly interested in this improvement and are currently developing user interface and operational concepts for execution stage automation.

Finally, it is interesting to note that in their analyses of aviation accidents, Wiegmann and Shappell found that the frequency of errors increased monotonically with information processing stage. The relative frequency of errors across these information processing stages is remarkably similar to what we found in our spacecraft environment, suggesting that error-reducing technologies designed for aviation may be useful for space applications, and vice-versa.

References

[1] J.W. McCandless, R.S. McCann, and B.R. Hilty, "Upgrades to the caution and warning system of the space shuttle," Proc. Of the Human Factors and Ergonomics Society 47th Annual Meeting, Denver, CO, pp. 16-20, October 2003.

[2] T.B. Sheridan, *Telerobotics, Automation, and Human Supervisory Control*, M.I.T. Press, Cambridge, MA, 1992.

[3] J.W. McCandless and R.S. McCann, "Development of new displays for the cockpit of the space shuttle," IBM Make Information Technology Easy Conference, San Jose, CA, 2002.

[4] J. Rasmussen, "Models of mental strategies in process plant diagnosis" in *Human Detection and Diagnosis of System Failures*, J. Rasmussen and W.B. Rouse, Eds. Plenum, 1981.

[5] N.B. Sarter and H.M. Alexander, "Error types and related error detection mechanisms in the aviation domain: An analysis of Aviation Safety Reporting System incident reports," *The International Journal of Aviation Psychology*, Vol 10, No. 2, pp. 189-206, 2000.

[6] J.W. Senders and N.P. Moray, *Human Error Cause, Prediction, and Reduction*, Erlbaum, 1991.

[7] D.A. Wiegmann and S.A. Shappell, "Human factors analysis of post-accident data: Applying theoretical taxonomies of human error," *The International Journal of Aviation Psychology*, Vol 7, pp. 67-81, 1997.

[8] C. Wickens and J. Flach, "Information processing" in *Human Factors in Aviation*, E. Wiener & D. Nagel, Eds., New York: Academic, 1988.